

(19)



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

**EP 0 780 777 A1**

(12)

## EUROPEAN PATENT APPLICATION

(43) Date of publication:

25.06.1997 Bulletin 1997/26

(51) Int Cl.<sup>6</sup>: **G06F 17/30, G11B 27/10**

(21) Application number: 96309056.8

(22) Date of filing: 12.12.1996

(84) Designated Contracting States:  
**DE FR GB**

(30) Priority: 21.12.1995 US 576106

(71) Applicant: **Hewlett-Packard Company**  
**Palo Alto, California 94304 (US)**

(72) Inventors:

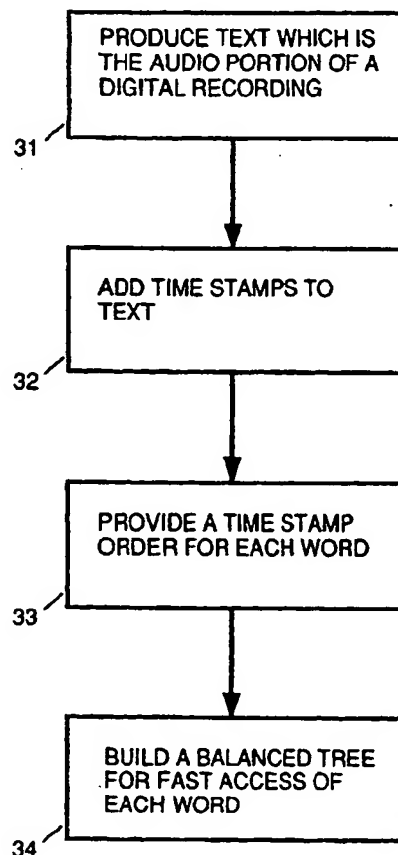
- **Orsolini, Garry S.**  
**Fair Oaks, California 95628 (US)**

- **Bowers, Richard D.**  
**Roseville, California 95747 (US)**

(74) Representative: **Jehan, Robert et al**  
**Williams, Powell & Associates,**  
**34 Tavistock Street**  
**London WC2E 7PB (GB)**

### (54) Indexing of recordings

(57) A recording is indexed by keywords. In order to perform the indexing, an audio portion (12) of the recording is transcribed (31) to produce text in a text file. A time stamp (32) is associated with each word in the text. Each time stamp (32) indicates a time in the recording at which occurs an associated word. Once a recording has been indexed, the recording may be searched along with other recordings. For example, in response to a user choosing a keyword (46), a text file for each recording is searched for occurrences of the keyword (46). At the conclusion of the search, each recording which includes an occurrence of the keyword is listed (42). When a user selects (42) a first recording and a particular occurrence of the keyword (46), the first recording is played starting slightly before a time corresponding to a first time stamp associated with the particular occurrence of the keyword in the first recording. In response to control sequences, prior and next occurrences of the keyword (42) can be observed in one or multiple recordings.



EP 0 780 777 A1

FIGURE 2

searching of recordings, improving over other currently available schemes to index recordings.

An embodiment of the present invention is described below, by way of example only, with reference to the accompanying drawings, in which:

Figure 1 illustrates steps taken to allow keyword indexing of digital recordings in accordance with the preferred embodiment.

Figure 2 is a flowchart which shows steps by which text for a digital recording is keyword indexed in accordance with the preferred embodiment.

Figure 3 and Figure 4 show computing displays which illustrate the preparation of a data base used for keyword indexing of digital recordings in accordance with the preferred embodiment.

Figure 5 shows a computing display used for keyword index searches of a video library in accordance with the preferred embodiment.

Figure 6 shows a computing display used for keyword index searches of a video library in accordance with an alternate embodiment.

Figure 1 illustrates steps taken to allow keyword indexing of digital recordings. A recording source 11 is digitized and compressed to produce digitized recording file 13. Recording source 11 is, for example, an audio recording or an audio-video recording. When recording source 11 is an audio-video recording, data in digitized recording file 13 is, for example, stored in MPEG-1 format. Digitized recording file 13 may be produced from analog recording source 11 using, for example, OptiVideo MPEG 1 Encoder available from OptiVision, having a business address of 3450 Hillview Ave., Palo Alto, CA 94304.

In addition, the audio portion of recording source 11 is transcribed to produce a text file 12 which includes the text. The transcription may be performed manually. Alternately, the audio portion of recording source 11 may be transcribed directly from recording source 11 or digitized recording file 13 using computerized speech recognition technology such as DragonDictate for Windows available from Dragon Systems, Inc., having a business address of 320 Nevada Street, Newton, MA 02160. Text file 12 and digitized recording file 13 are then made available to a computer system 14.

Figure 2 is a flowchart which shows steps by which text for a digitized recording file 13 is keyword indexed. In a step 31, text is produced which is the audio portion of digitized recording file 13. This text is a result of the transcription described above.

Figure 3 illustrates the result of the transcription process. Figure 3 shows a window 23 in a computer screen 21. Within window 23 is the transcribed text of the audio portion of recording file 13.

In a step 32, shown in Figure 2, time stamps associated with words in the text are added to the transcribed text. In the preferred embodiment, the time stamps are in milliseconds and indicate elapse of time relative to the starting point of the digital recording within recording file

13.

Placement of time stamps may be performed, for example, with the help of an operator utilizing, on computer 14 (shown in Figure 1), software specifically designed to add time stamps. For example, the recording is played by computer 14. For an audio-video recording, a window 22 in computer screen 21, as shown in Figure 3, may be added in which the audio-video recording is played. The operator of computer 14, using cursor 24, selects words as they are spoken in the recording played by computer 14. Whenever the operator selects with cursor 24 a word from the text in window 23, the software running on computer 14 time stamps the word with the current time duration which represents the elapse of time relative to the starting point of the digital recording.

Figure 4 further illustrates this process. In Figure 4, time stamps TS1, TS2 and TS3 have been added to text 23 by an operator as described above. Source code for software which implements the time stamp feature discussed above for audio-video recordings will be apparent to the skilled person. Alternately, step 32, shown in Figure 2, may be automated so that speech recognition technology is used to trigger the placement of time stamps within text 23.

After the time stamps have been added to text 23, in a step 33 shown in Figure 2, every word of text 23 is assigned a time stamp. For words which were not assigned a time stamp in step 32, interpolation is used to determine an appropriate time stamp.

For example, Table 1 below shows a portion of text 23 after the completion of step 32.

Table 1

Once::11 upon a time::20 there was a boy::28 named Fred. He went::35 to the forest::44. . . .

In the example given in Table 1, the word "Once" was spoken at 11 milliseconds from the beginning of the audio track of the digital recording. The word "time" was spoken at 20 milliseconds from the beginning of the audio track of the digital recording. The word "boy" was spoken at 28 milliseconds from the beginning of the audio track of the digital recording. The word "went" was spoken at 35 milliseconds from the beginning of the audio track of the digital recording. The word "forest" was spoken at 44 milliseconds from the beginning of the audio track of the digital recording.

In order to assign time stamps to the remainder of the words, interpolation is used. For example, nine milliseconds elapsed between the word "Once" and the word "time". There are two words, "upon", and "a", which occur between "Once" and "time". As a result of the interpolation, the words "upon", and "a" are assigned time stamps of 14 milliseconds and 17 milliseconds, respectively. This is done so that there is allocated three milliseconds between the occurrence of the word "Once" and the word "upon"; there is allocated three millisec-

rence of the keyword, go back to the last occurrence of the keyword, continue playing and so on. The interface also includes a "cancel" button 49.

In addition to searching on one or more keywords connected by Boolean variables, the balanced tree formed in step 34 (shown in Figure 2) may also be searched using concept based searching techniques, for example using Metamorph available from Thunderstone Software-EPI, Inc. having a business address of 11115 Edgewater Drive, Cleveland, Ohio 44102.

The foregoing discussion discloses and describes merely exemplary methods and embodiments.

The disclosures in United States patent application no. 08/576,106, from which this application claims priority, and in the abstract accompanying this application are incorporated herein by reference. The US parent application also includes examples of the source codes mentioned herein.

## Claims

1. A method of indexing a recording comprising the steps of:
  - (a) transcribing (31) an audio portion of the recording to produce text in a text file; and,
  - (b) providing (32) for each of a set of words in the text, a time stamp which indicates a time in the recording at which each word in the set of words occurs.
2. A method as in claim 1 wherein step (a) is accomplished manually by a transcriber or with the use of speech recognition technology.
3. A method as in claim 2 wherein when step (a) is accomplished with the use of speech recognition technology, steps (a) and (b) are performed simultaneously.
4. A method as in claim 1, 2 or 3, wherein step (b) includes the substeps of:
  - (b.1) providing for each of a subset of the set of words in the text, a time stamp which indicates a time in the recording at which each word in the subset of the set of words occurs; and,
  - (b.2) for a remainder of the set of words which are not in the subset of the set of words, using interpolation to provide a time stamp which indicates a time in the recording at which each word in the remainder of the set of words occurs.
5. A method as in claim 4 wherein the recording is an audio-video recording and wherein substep (b.1) includes the substeps of:
  - (b.1.1) displaying the text in a first window (23) of a computer display;
  - (b.1.2) playing a video portion of the recording in a second window (22) of the computer display; and,
  - (b.1.3) upon an operator selecting a selected word of the text in the first window, adding a time stamp (TS1...) to the text file which indicates an elapsed time from a beginning of the recording until selection by the operator of the selected word.
6. A method as in any preceding claim, comprising the step of:
  - (c) arranging the set of words and associated time stamps into a balanced tree based on occurrences of each word in the set of words.
7. A method of accessing selections within a plurality of recordings, comprising the steps of:
  - (a) in response to a user choosing a keyword, searching a plurality of text files for occurrences of the keyword, wherein text files are associated with recordings so that for each of the plurality of recordings, one text file from the plurality of text files includes a text of an audio portion of the recording, each word in each text file being associated with a time stamp (TS1...) which indicates an approximate location in an associated recording of an occurrence of the word;
  - (b) listing (44) recordings which include an occurrence of the keyword; and,
  - (c) upon a user selecting a first recording and a particular occurrence of the keyword, playing the first recording starting slightly before a time corresponding to a first time stamp associated with the particular occurrence of the keyword in the first recording.
8. A method as in claim 7 wherein in step (c) upon a user selecting the first recording, a first-in-time occurrence of the keyword within the first recording is automatically selected as the particular occurrence of the keyword.
9. A method as in claim 7 or 8 wherein step (b) includes the substeps of:
  - (b.1) listing in a first window the recordings which include an occurrence of the keyword;
  - (b.2) highlighting one of the recordings from the recordings listed in the first window; and,
  - (b.3) listing each of the occurrences of the keyword within the recording highlighted in substep (b.2).
10. A system for accessing selections within a plurality

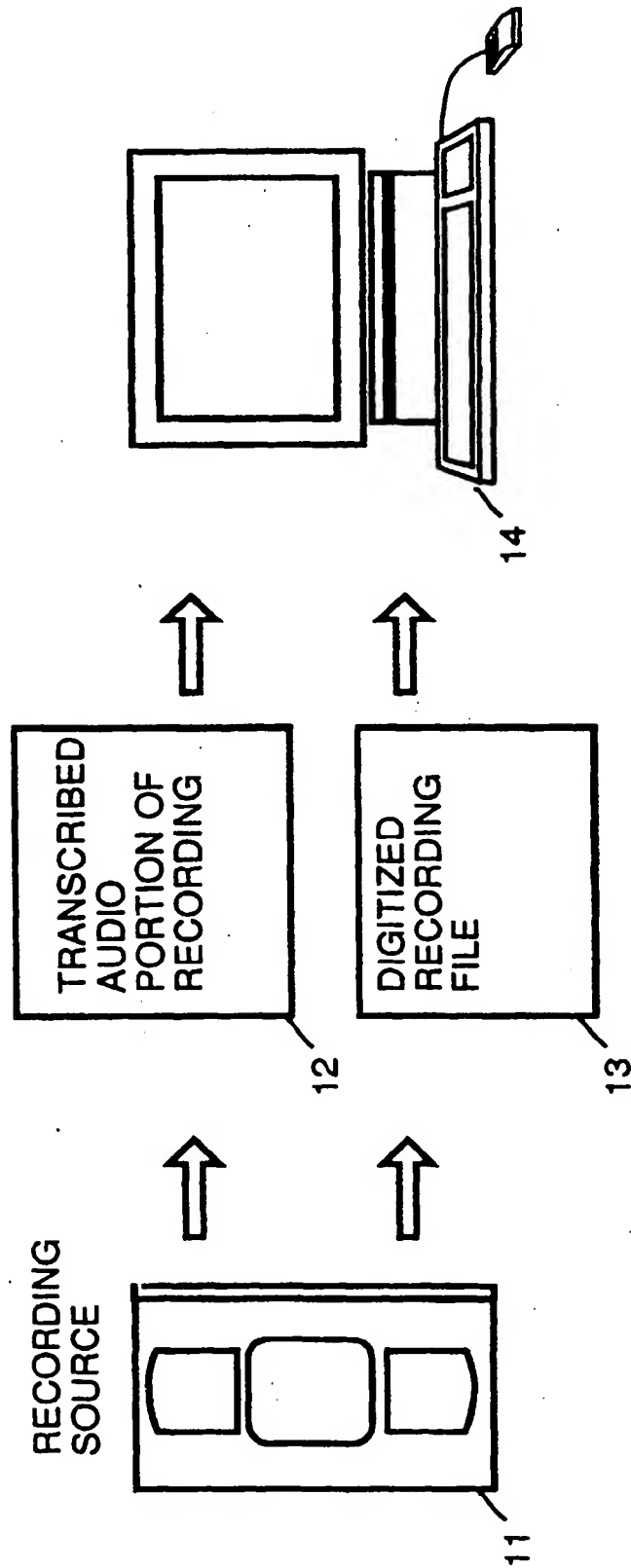


FIGURE 1

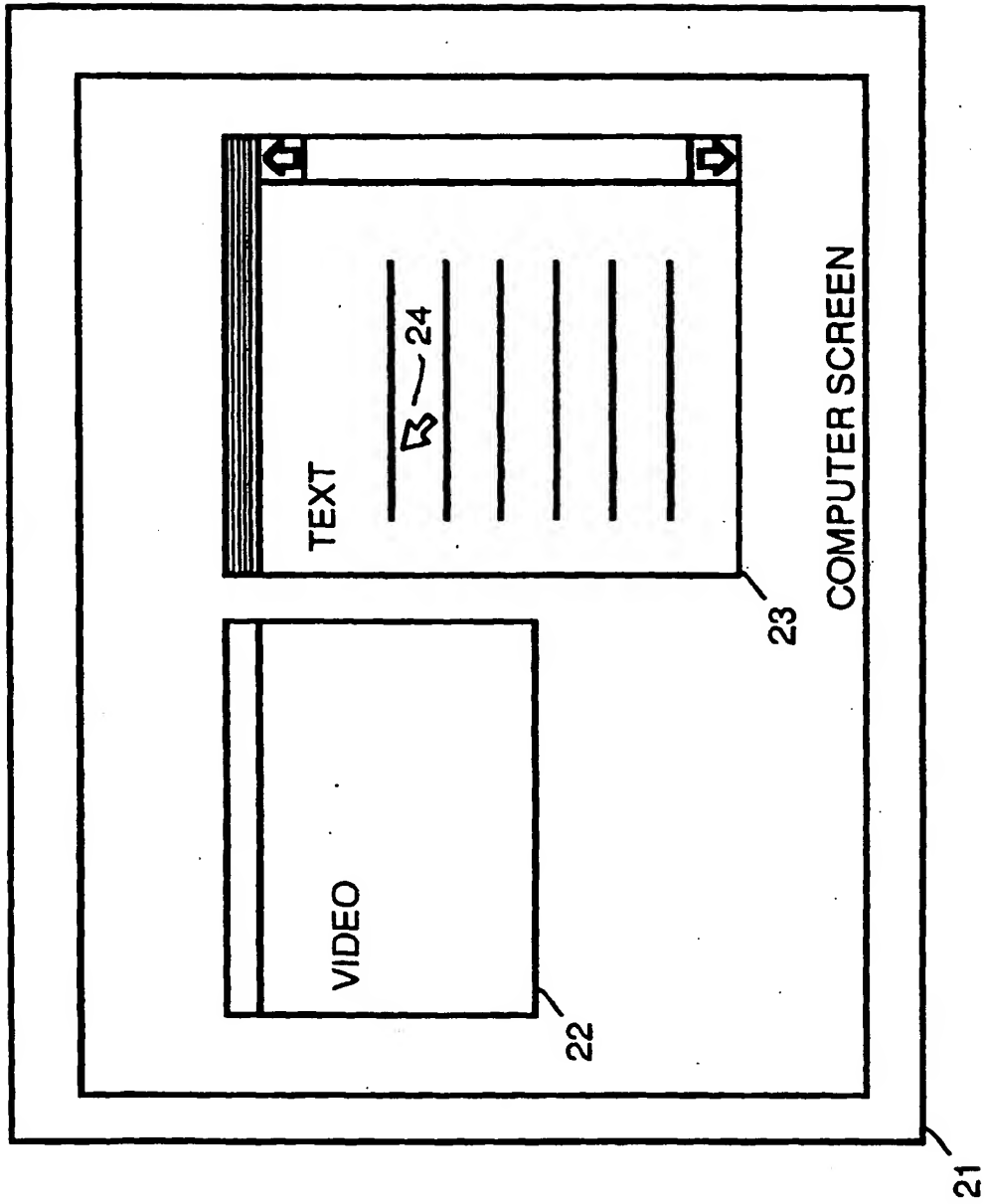


FIGURE 3

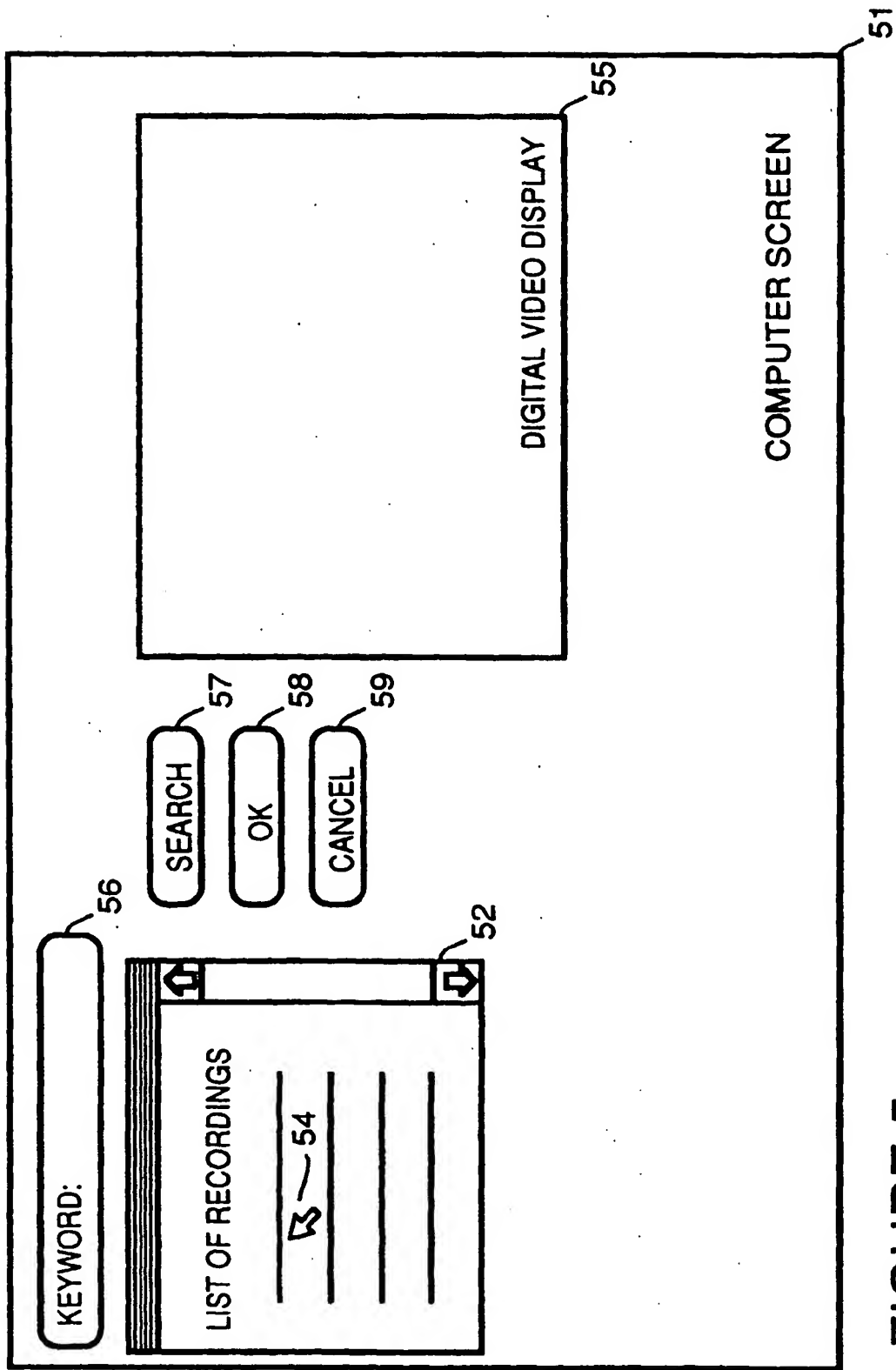


FIGURE 5



European Patent  
Office

# EUROPEAN SEARCH REPORT

Application Number  
EP 96 30 9056

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	EP 0 507 743 A (STENOGRAPH CORP) 7 October 1992 * page 2, line 23 - line 48 * * page 5, line 25 - line 53; figures 5,6 * * page 6, line 14 - line 36 *	1-4,7-13	G06F17/30 G11B27/10
X	EP 0 649 144 A (IBM) 19 April 1995 * column 4, line 5 - column 6, line 14; figures 1,2 *	1-5 6	
Y	* column 13, line 37 - column 14, line 8; figure 5 *	7,10	
A	* column 6, line 18 - line 46; figure 3 *		
X	US 5 136 655 A (BRONSON BARRY S) 4 August 1992 * column 2, line 67 - column 3, line 42; figure 1 *	1-3	
Y	PATENT ABSTRACTS OF JAPAN vol. 013, no. 590 (P-984), 26 December 1989 & JP 01 253027 A (NIPPON TELEGR & TELEPH CORP), 9 October 1989, * abstract *	6	TECHNICAL FIELDS SEARCHED (Int.Cl.6)  G06F G11B
X	WO 92 11634 A (ARDIS PATRICK M ; MARKOVICH MARKO R (US); THOMPSON KEVIN W (US)) 9 July 1992 * abstract; figure 3 * * page 9, line 21 - page 11, line 22 *	1-3,10	
X	IBM TECHNICAL DISCLOSURE BULLETIN, vol. 33, no. 10A, 1 March 1991, page 295/296 XP000110048 "CORRELATING AUDIO AND MOVING-IMAGE TRACKS" * the whole document *	1,2,10	
The present search report has been drawn up for all claims			
Place of search BERLIN		Date of completion of the search 1 April 1997	Examiner Deane, E
CATEGORY OF CITED DOCUMENTS X: particularly relevant if taken alone Y: particularly relevant if combined with another document of the same category A: technological background O: non-written disclosure P: intermediate document		T: theory or principle underlying the invention. E: earlier patent document, but published on, or after the filing date D: document cited in the application I: document cited for other reasons &: member of the same patent family, corresponding document	

EPO FORM 150 (01.92) (P/4001)